运用区块链技术构建数字化古籍管理体系模型的研究*

高利[1] 王春艳[1] 高心丹[2]
[1] 东北林业大学图书馆 哈尔滨 150040 [2] 东北林业大学信息与计算机工程学院 哈尔滨 150040

摘要：［目的/意义］从区块链技术角度出发，针对当前古籍保护与利用领域存在的问题，提出数字化古籍管理区块链体系模型的解决方案，实现古籍管理创新模式。［方法/过程］首先，从构建和实现两个方面设计基于区块链去中心化的P2P体系架构。通过节点发起交易、构造区块、合法性验证实现全网共识链接区块。其次，建立信息存储机制、智能合约协议算法，自动执行保证模型的高效、透明，实现数字化古籍版权登记、流转利用信息不可篡改、可追溯。最后，共识机制和IPFS协议系统分布式存储机制解决古籍数据存储和安全，同时设立合理的激励机制，实现古籍数据的有效管理。［结果/结论］通过该模型的建立摒弃传统古籍数据库中心化管理模式，以区块+链的形式实现古籍保护利用全过程可追溯、不可篡改，从而实现古籍保护的技术创新。

关键词：区块链 数字化古籍 智能合约 IPFS
分类号：G250
DOI：10.13266/j.issn.0252-3116.2019.03.011

1 引言

2017年11月8日习近平主席会见美国总统特朗普，在参观故宫时曾向其介绍说："中国是具有5 000多年历史，虽然埃及历史更长一些，但是，文化没有断过流、始终传承下来的只有中国"，中华文化的传承靠的就是浩如烟海的古代典籍和文献。钱振东在《书厄述要》中写道："文化之于国家，犹精神之于形骸。典籍者，又文化之所赖以传焉者也。"[1]中国历代古籍记载和证实了中华文明绵延不绝的历史，传承着中华优秀文化所蕴含的人文精神和道德规范。随着世代更迭、岁月沧桑，古籍历经浩劫磨难，留存至今的古籍善本价值益发凸显，各收藏单位均视若珍宝[2]。数字化古籍是为了保护和利用古籍这一目的，是我国经典古籍整理的新方式，是我国传统文化传播的新途径，是我国精神文明建设的新标志[3]。我国自"十一五"就提出启动编撰以《中华大典》为代表的古籍全书数字化国家重大出版工程。近年来随着国家有关部门对数字化转型升级给予大力的支持，数字化古籍资源也焕发出新的生命力[4]。但数字化古籍在快速发展的同时也出现了诸如所有权、流通利用、数据安全等亟待引起重视的问题。

区块链作为数字加密货币比特币（BitCoin，BTC）的底层技术，近两年来受到国际社会的广泛关注[5]。我们对2015年至2018年5月间相关文献进行了归纳总结，如表1所示：

表1 以"区块链"和"Blockchain"为主题词文献检索统计

数据库名	2015年（篇）	2016年（篇）	2017年（篇）	2018年5月（篇）
SCIE（包括OA资源）	6	47	227	133
EI	31	156	563	229
CNKI	35	986	2 222	1 791
合计	72	1 189	3 012	2 153

通过数据库检索，区块链相关研究自2015年逐步展开，2017年开始研究数量持续高涨。按EI受控词汇分析，区块链研究从电子货币、经济学、分布式计算、网络安全逐渐转移到电子货币、数据安全、分布式数据隐私、物联网等领域，其中发文量集中在数字金融和数据安全领域。在CNKI检索区块链相关文献，使用组合词进行检索，结果是"区块链+版权"共70篇文献；"区块链+数据安全"24篇；"区块链+数据保护"10篇。近两年学术领域运用区块链技术，针对数字版权、数据分布式存储、安全共享、防伪溯源等领域开展了持

* 本文系黑龙江省艺术规划项目"区块链技术在知识共享服务领域应用的研究"（项目编号：2018D013）研究成果之一。
作者简介：高利（ORCID：0000-0001-7693-5003），馆员，博士；王春艳（ORCID：0000-0001-7451-1864），馆员，硕士；高心丹（ORCID：0000-0001-8008-7828），副教授，博士，硕士生导师，通讯作者，E-mail：124239774@qq.com。
收稿日期：2018-06-03 修回日期：2018-08-09 本文起止页码：80-89 本文责任编辑：易飞

续性的研究。贾引狮[6]从当前版权登记存在的问题，用辩证思维运用区块链技术解决网络版权交易。郑阳与杜荣[7]在分析国内外区块链数字资产管理平台基础上，建立区块链数字资产一站式服务。李悦等[8]提出了数字作品版权(DCI)管理模型，通过版权登记解决版权问题。郝琨等[9]针对传统分布式存储存在的问题，构建区块链去中心化分布式存储模型，保证数据安全及可追溯性。吴振铨等[10]运用联盟区块链存储系统，解决中心化数据库易引起恶意攻击发生单点失效问题。高灵超等[11]结合数据共享过程标识不唯一、易窃取或篡改问题，设计了区块链数据共享 P2P 安全可信网络体系。

2 区块链数字化古籍管理体系适用性及研究意义

2.1 数字化古籍管理应用区块链的适用性

2.1.1 主体适用　数字化是古籍保护和利用的有效手段，数字化古籍管理体系需要众多的参与主体来协助与支撑，包括拥有古籍文献的图书馆、博物馆、收藏机构或个人、数字化厂商、数据库公司等。其参与主体呈现多元化特征，各主体形成多个具有分散性和自治性的古籍数据中心。而区块链系统通常采用对等式的 P2P 网络，为去中心化的分布式组织。各节点地位对等且所有节点通过加密算法，提供工作量证明进行区块验证，以扁平式拓扑结构相互联通，不存在中心化的特殊节点和层级结构。因此数字化古籍管理体系的主体特征与区块链去中心化、分布式存储等特点高度契合。

2.1.2 客体适用　区块链技术的核心就是利用加密链式区块结构对数据区块进行验证与存储，所有用户节点通过共识算法对数据生成、更新并加盖时间戳。当前区块链应用领域可以概括为数字货币、金融交易、产品溯源、数据存储、数据鉴证、数字资产管理等，能够实现对各类数字作品的产权等无形资产进行登记、确权、授权及实时监控。而古籍管理的对象为数字化古籍，从其收集整理、数字加工、标引分类、共享流通到授权使用，整个流转过程通常可以抽象为数据资产的形成与交易过程。因此数字化古籍管理的客体构成与区块链数字资产管理应用场景存在高度契合。

2.1.3 功能适用　当前数字化古籍管理在给研究者和读者带来便利的同时，也存在诸如下列的一些问题，如：①数字化古籍侵权案例增多，同行间互相剽窃、盗版现象时有发生；②数字化古籍内容把关缺失，为迎合市场需要对古籍内容随意进行增删修改；③缺少激励机制，一些拥有古籍文献的公共图书馆、博物馆或是个人收藏者对于数字化古籍并不热衷，导致一些传统古籍数字化难度增大；④数字化古籍都是以中心化的古籍数据库管理形式出现，其中全文影像数据库和图文对照数据库更有利于古籍的保护与利用，但也会使存储量增加，数据安全方面也存在问题。上述这些问题的产生，究其主要原因是各参与主体之间信任基础薄弱，以及古籍数据库中心化管理，并且缺少有效激励措施而产生的。区块链技术特性是所有节点参与验证、不可篡改、可追溯及形成共识的激励机制，这与构建数字化古籍管理体系、增强古籍保护与利用的功能需求相匹配。

2.2 本研究的意义

本文将区块链技术引入数字化古籍管理领域，设计基于区块链的数字化古籍管理模型。通过区块链自身的不可篡改、时间戳、可追溯、分布式存储等特点[12]，为数字化古籍管理提供版权保护、流通共享、数据安全存储等方面的全链上服务。古籍数据作为区块是由参与者构造、验证、链接，网络中发生的每一次交易(操作)信息都会永久保存在区块链中，且不可更改。这就使用户可以对数字化古籍版权进行登记，并且得到全网共识确认，同时数字化古籍被下载、拷贝、浏览等过程都有记录可查，保证数字化古籍版权的权威性和溯源性。

针对当前古籍数据库存在的侵权、篡改内容、缺少激励措施、数据存储安全方面，提出运用区块链技术构建数字化古籍管理体系的创新方法，以实现所有节点共同维护、提高数字化古籍数据质量、安全存储、解决知识产权纠纷、提高共享意愿及古籍利用率；摒弃传统的古籍数据库中心化管理方式，以区块＋链的形式完成数字化古籍的数字化－存储－共享交易全部管理流程。

3 构建模型的技术基础

3.1 区块链基本原理

区块链技术(blockchain technology，BT)也被称为分布式账本技术，特点就是去中心化、不可篡改、可追溯、公开透明，让每个人都可以参与数据库记录。区块链的主要概念包括：①交易(transaction)：操作一次，就会导致账本状态改变一次，添加一条记录；②区块(block)：记录一段时间内发生的交易和状态结果，是对当前账本状态的一次共识；③链(chain)：由一个个

区块按照发生顺序串联而成,是整个状态变化的日志记录。在区块链中,数据以区块的方式永久储存。区块链作为一个状态机,会对网络中每次状态变化进行确认、记录生成区块,并按时间顺序连接成链,每一个区块记录了创建期间发生的所有交易信息[13]。

区块的数据结构一般分为区块头和区块体两个部分[14]。区块头记录信息 H,包括:版本号(Version)、前一区块 Hash 值、默克尔根(Merkle root)、时间戳(Time)、难度值(Target_bits)、随机数(Nonce)。大小为固定 80 个字节。区块体记录区块产生对应的交易信息 T 和包含的其他所有信息 U。

区块 B 的形式化定义如式(1)所示:

$$B = (B_H, B_T, B_U) \quad 式(1)$$

区块链中区块体记录区块生成这段时间里所有的交易信息,大小不固定,与区块体相比,虽然区块头比区块体小很多,但大部分功能都由区块头实现。

3.2　智能合约

智能合约(Smart Contract)是一组预设条件的计算机代码,由学者尼克. 萨博(Nick Szabo)于 1995 年提出。智能合约作为区块链上的触发器,可准确自动执行,无人工干预[15]。

数字化古籍管理区块链网络中智能合约包括事物处理和保存机制,以及一个完备的状态机,用于接受和处理各种智能合约,并且事务的保存和状态处理都在区块链上完成。这个过程就是将预设规则代码封装在区块中,根据事件描述信息设定触发条件,经过数字化古籍管理区块链网络节点,进行验证,然后链接同步。同时区块链网络实时监控区块中的部署合约,触发智能合约进行状态机判断。如果自动状态机中某个或某几个动作的触发条件满足,则由状态机根据预设信息选择合约动作自动执行合约。

4　模型架构设计方案

基于区块链技术建立数字化古籍管理网络,可以实现去中心化的共识信任机制和分布式存储,改变传统数字化古籍保护与利用活动当中的各方关系。构建 P2P 体系架构,以区块链作为核心存储机制,通过创建交易、构造区块、验证区块合法性、链接区块,来实现数字化古籍版权登记、流通、数据安全的管理模式。具体数字化古籍管理区块链网络体系架构的模型设计如图 1 所示:

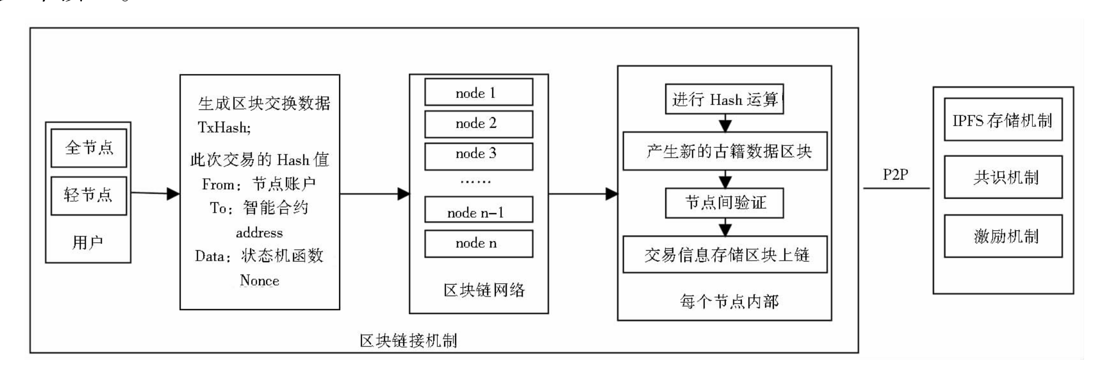


图 1　数字化古籍管理区块链模型架构

从该模型架构中可以看出数字化古籍管理模型包括用户节点构建区块上链机制和区块链 P2P 网络实现机制两个部分。

4.1　用户节点构建区块上链机制

4.1.1　*交易发起*　用户节点是区块链网络的客户端,是连接到区块链网络的设备终端。用户节点在客户端进行的每一次操作请求,都可以看作是发起一次交易,产生交易数据。区块链节点会将交易数据提交区块链中部署的智能合约状态机,触发远程过程调用协议(Remote Procedure Call Protocol,RPC)接口,执行状态机中预先部署的智能合约算法,实现数字化古籍版权登记、查证、交易的一系列功能。

在该模型中设计的智能合约算法其自动执行的工作请求包括 3 类:①数字化古籍版权登记请求,记作 T_α;②数字化古籍使用授权请求,记作 T_β;③数字化古籍版权查证请求,记作 T_γ;例如,执行版权登记操作具体执行流程为:数字化古籍管理区块链网络中,版权用户对自己的数字化古籍进行版权登记,节点会向区块链网络发起一个 T_α 请求,部署在区块链网络中的智能合约算法将自动执行一次 T_γ 请求,遍历区块链网络后,如果 T_α 为真,那么发起的这次新交易会向全网广播,智能合约算法会将数字化古籍版权信息写入新区块。新区块的区块头会将本次交易信息进行记录,并加盖时间戳,这样区块链上的这一古籍版权信息数据

将不可篡改，并且此数据的每次操作信息、流向数据都会按时间顺序以添加记录的形式记录到区块体中，T_β、T_γ 的交易发起流程也是如此。这样就能够为数字化古籍的版权归属提供权威证明。

4.1.2 交易处理 区块链网络中的交易处理过程也就是产生新区块的过程，能否完成交易处理、产生新区块，是由节点和区块链网络中的共识机制共同决定的。节点决定着谁有出块的资格，共识机制则决定着这些有出块资格的节点最终哪个可以出块。

区块链的共识机制是所有的节点都通过计算 Hash 来实现工作量证明(Proof of Work，POW)。假如现在有顺序产生的数据块 A、B、C，计算 Hash 头的形式化定义如式(2)所示：

$$H0 = Hash(A \parallel Nonce0)$$
$$H1 = Hash(B \parallel H0 \parallel Nonce1) \quad 式(2)$$
$$H2 = Hash(C \parallel H1 \parallel Nonce2)$$

式(2)中，由于密码学 Hash( )函数的单向性：Hash(x) = y，通过 y 很难找到 x；‖ 表示链接；Nonce 是随机数。当节点接收到用户发起的交易请求后，数字化古籍管理区块链网络中形成共识机制，会在本地节点进行数学运算来获取记账权，然后计算当前最新的 Hash 的头部，当成功计算出了一个符合要求的 Hash 后，新的数字化古籍版权区块产生，新区块会在全网节点之间广播，每次达成共识需要全网共同参与运算，当超过 50% 节点确认该区块有效后，本次版权信息才会被写入区块链中存储[16]。整个区块链交易处理过程由智能合约通过代码自动执行，不受干预，确保整个流程公开透明。

4.2 区块链 P2P 网络实现机制

数字化古籍区块链网络是由所有参与者各个节点共同构建的分布式 P2P 网络。区块链网络中区块存储的核心内容是交易信息，区块链形成过程就是各节点以最快的运算速度对区块数据的有效性达成共识的过程。

4.2.1 交易信息存储共识机制的优化 目前区块链共识机制主要是采用工作量证明 POW 机制。是节点在处理交易数据的同时，运用 SHA256 算法通过不断 Hash 计算，计算出随机数 Nonce。当全网有一位节点计算出 Nonce 时，他就会把自己打包的区块公布出去，其他节点会进行验证，验证通过后就会一致性认为这个区块链接到了区块链上，然后继续进行下一个区块的构造和 Hash 计算[17]。在这个过程中会有分叉块的产生，并且需要等待多个确认，这种简单暴力的方法能保证整个区块链系统的合法性。但这种共识机制在数字化古籍管理的交易信息存储体系中应用上存在明显的不足。为此在数字化古籍管理区块链网络中，为使共识区块自身计算最大化收益的单一行为，以及保障交易信息数据存储安全性和有效性之间的关系，本文采用节点验证组合及 POW 收益集合的共识机制达成全网共识[18]。

在构建新区块的 Hash 计算过程中，当所有节点都有待提交验证信息时，节点为了自身收益，每一区块都不会改变自己对别的区块的验证结果[19]。在交易信息存储系统中的数学形式描述为：$B = \{B_1, B_2, \cdots, B_n\}$，为节点集合。假如：节点 B_i 运算后构造信息的区块得到其他节点验证组合和 POW 收益，可用集合 $G_i = \{V_{il}, \cdots, V_{in}: u_i\}$ 来表示。那么节点 B_i 构造后信息区块组成的各节点验证组合($V_{il}, \cdots, V_{in}$)中，任何一个参与验证的节点 B_j 提交给节点 B_i 的验证结果为 V_{ij}，并且满足式(3)至式(5)：

$$V_{ij} = \begin{cases} 1，经过节点 B_j 验证，节点 B_i 提交的区块为真 \\ -1，经过节点 B_j 验证，节点 B_i 提交的区块为假 \end{cases} \quad 式(3)$$

$$B_j \xrightarrow{V_{ij}=1} B_i : u_i = u_i + 1 \quad 式(4)$$

$$B_j \xrightarrow{V_{ij}=0} B_i : u_i = u_i - 1 \quad 式(5)$$

那么取得这一轮交易信息区块记账权可以分为以下 3 种情况：

(1)如果这一轮交易信息区块 Hash 运算还没有结束时，当最先出现 $B_i \in B$，并且 $u_i(V_{il}, \cdots, V_{ij}, \cdots, V_{in}) = n$，那么 B_i 就是这一轮 Hash 运算的最佳区块，也就是能通过区块链网络所有节点验证的区块。

(2)如果这一轮交易信息区块 Hash 运算已经结束，当 $\exists B_i, \forall B_i \in B$，并且 $n > u_i(V_{il}, \cdots, V_{ij}, \cdots, V_{in}) > u_j(V_{il}, \cdots, V_{ij}, \cdots, V_{in})$，那么 B_i 就是这一轮 Hash 运算的最佳区块，也就是能通过区块链网络所有节点验证的区块。

(3)如果这一轮交易信息区块 Hash 运算已经结束，当 $\exists B_i, B_j, \forall B_k \in B$，并且 $n > u_i(V_{il}, \cdots, V_{ij}, \cdots, V_{in}) = u_j(V_{jl}, \cdots, V_{ji}, \cdots, V_{jn}) > u_k(V_{kl}, \cdots, V_{kj}, \cdots, V_{kn})$，那么从 B_i 和 B_j 当中选最先达到 u_i 当前值的区块就是这一轮 Hash 运算的最佳区块，也就是能通过区块链网络所有节点验证的区块。

4.2.2 执行共识机制构建区块与链接 数字化古籍管理区块链网络中的任何一个节点，如果要生产一个新交易信息区块并写入区块链，需要 3 个要素，即区

块、难度值、工作量证明函数。构造新区块过程当中通过同步最新区块链来取得下一区块的记账权，对这期间网络中没能确认的交易进行收集 $T=\{T_1, T_2, \cdots, T_n\}$。数字化古籍交易区块难度值决定产生新区块需要进行 Hash 运算量，新区块头的难度值 H_d 用来设定下一步的挖矿目标 Target。在交易执行和验证过程中，对交易的数字签名和合法性进行验证，验证为真的交易会 SHA256(SHA256( ))计算，计算出交易信息的 Hash 值，记录在区块头的 Merkle 根。此外，会根据数字化古籍区块所包含的重要信息进行计算和收集，形成交易信息数据列表，记录在区块体的 Transaction Infor 处，数字化古籍区块信息如图 2 所示：

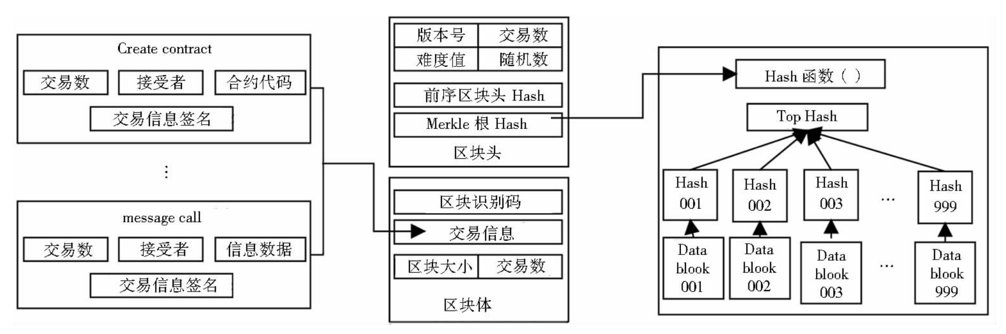


图 2 数字化古籍区块信息

由此，构造出不包含随机数 Nonce 的交易信息区块 B_i，再执行共识算法 POW 构建数字化古籍区块 B。根据 B_i 区块头的难度值 H_d，定义这一区块的难度系数为 $Target$，$Target=2^{256}/H_d$，会相应地生成一个随机数 Nonce，然后通过 POW 函数运算得到 $FinalHash$ 值，如果 $FinalHash<Target$，那么构建数字化古籍管理区块 B 成功。该构建区块过程形式化表示如式(6)所示：

$$F \leq 2^{256}/H_d \parallel F=PoW(H_i, T_x, nonce, datatorrent) \quad 式(6)$$

其中，F 表示符合条件的 $FinalHash$ 值，POW 表示工作量共识算法函数，H_i 为 B_i 的区块头，T_x 为数字化古籍管理区块 B 中确认交易的 Hash 值，Nonce 为随机数，Datatorrent 为这一区块伪随机种子产生的数据集。具体构造区块流程如图 3 所示：

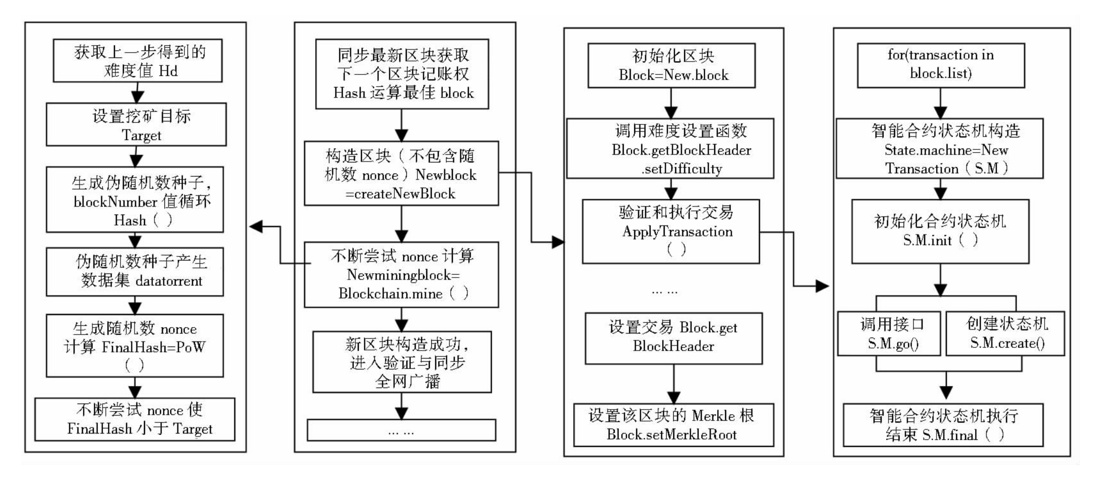


图 3 数字化古籍管理区块链网络节点构造区块流程

构造出新的数字化古籍区块后，在全网广播这一区块，其他的节点对区块进行验证，如果为真则同步新区块到数字化古籍管理区块链上。数字化古籍区块链接的形式化表示如式(7)所示：

$$C_{t+1}=S(\cdots\eta(\eta(C_t, T_0), T_1)\cdots) \quad 式(7)$$

式(7)中 $t+1$ 时刻产生的区块 B，其确认的交易有 n 次($T_1, T_2, \cdots, T_n$)；S 为区块 B 链接的数字化古籍管理区块链状态转换函数；η 为单次交易产生区块链状态转变函数；C_t 是在 t 时刻区块链状态；C_{t+1} 是添加了新区块 B 后的区块链状态；通过了全网的验证与链接，交易才算是被确认并且永久地记录到数字化古籍管理区块链上，永远无法更改。

交易信息存储算法如算法 1 所示：

算法 1：交易信息数据区块存储算法

Input：A set N of Nodes，newblockB，

```
Input:blockchainC
Input:T_end,the end of create block stop
procedure store(N,B,C,T_end)
    P = {T_1,T_2,…,T_n}
    createBlockhead H()
if time < T_end then
    foreach nonce do
        if nonce < = 2^256/H_d &&mixhash = H(mixhash)
           then
          if (nonce,mixhash) = PoW() then
            foreach n ∈ N do
              verify(B)
                if B = true then
                  C = addBlock(C,B)
                foreach n ∈ N do
                  distributeBlockchain(C^ˊ,n)
    else updateTime()
end procedure
```

4.2.3 数字化古籍动态交易信息区块存储体系结构

数字化古籍管理区块链网络的存储体系采用多级访问控制模式,节点在客户端发起交易,建立创世块到交易处理构造新区块,再到通过全网验证区块上链。这一整个流程就是数字化古籍的版权登记到查询再到其他节点使用数字化古籍请求,完成数字化古籍流转的过程。整个过程完全是在区块链上完成,是一个交易数据动态存储上链过程。由于区块链的不可篡改特性,区块间信息数据是动态添加的。动态数据对应的是数字化古籍版权和使用权转移,下面分析相邻区块进行数据交互,区块间通信交互情况如图4所示:

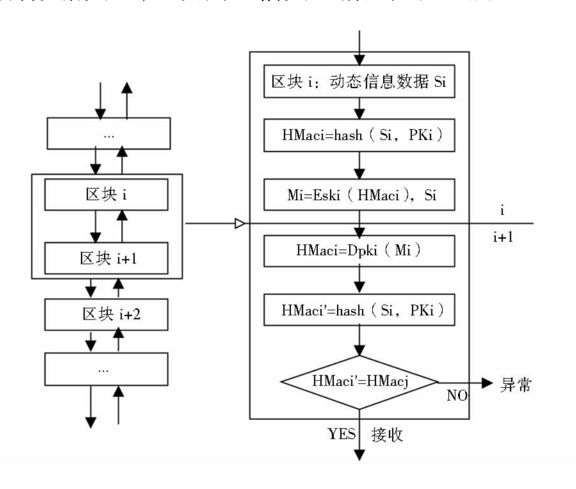


图4 相邻区块动态数据通信交互示意

区块链的基础是密码学,加密解密是由密钥实现。数字化古籍管理区块链网络为链上各节点生成密钥对(PK_i,SK_i),密钥对的作用就是相邻区块间信息数据通信,而且只允许相邻的区块进行通信。区块链上两个相邻的区块 B_i 和 B_{i+1}。假设区块 B_i 为发送方,区块 B_{i+1} 为接收方,区块 B_i 产生的动态交易信息数据编码为 S_i。通常为减少存储空间和提高计算速度,相邻区块进行通信时会采用二次散列迭代的方式,把发送方的公钥 PK_i 和信息编码 S_i 同时作为 Hash 函数的变量,可以得到作为特征值的 Hash 运算消息认证码(Hash based Message authentication code,HMac),数学计算公式如式(8)所示:

$$\mathrm{HMac}(PK_i,S_i) = \mathrm{H}(PK_i \oplus odata \parallel \mathrm{H}(PK_i \oplus idata \parallel S_i)) \quad 式(8)$$

式(8)中,PK_i 为发送方公钥,S_i 为所要发送的信息,H 为散列函数,⊕表示异或,‖ 表示链接,odata 与 idata 表示事先指定的字符串。

用发送方私钥对式(8)运算得到的消息验证码 HMac 进行签名。区块 B_i 将发送方签过名的消息验证码 HMac 和消息正文传送给区块 B_{i+1}。区块链各个节点的账号就是其公钥,节点账号使用自己的私钥会对已经通过验证的信息进行签名。新交易的创建和区块构建过程前文已介绍过,某一节点将该交易通过 P2P 网络进行全网广播,各节点区块会对该交易进行验证,并按照前文4.2.2节提出的共识机制选出各轮获得共识的区块,然后再次通过 P2P 网络进行全网广播。最后,以 Hash 方式将该交易信息区块链接到区块链上,同时将该链同步到区块链各个节点进行更新,从而完成动态交易信息的存储和交易数据的分布式记账。例如:在数字化古籍管理区块链中,新产生区块 B 如果用四元数组(s,t,l,c)表示,s 为区块序号,t 为区块链中不同区块交易信息类型,l 为新区块中这一次交易信息长度,c 为按照区块链标准建立的动态交易信息数据编码格式。假设数字化古籍管理区块链节点发起交易产生创世块,并且新生区块非空,那么区块有效验证算法如算法2所示:

算法2:新区块有效验证算法:

```
Input:newblockB,blockchainC
procedure validate_block(B,C)
    B←(x_c)
if B∧(C))then
    (s,t,l,c)←)
HMac (S_i,PK_i)
    HMac ←M(S_i,PK_i)
```

图书情报工作
第 63 卷 第 3 期 2019 年 2 月


```
if validblock(s,t,l,c) ∧ (HMac = = HMac))
    then
blockchainC←l ‖ h(tail(C))
    else
B = false
    end if
end if
    return(B)
end procedure
```

4.3 基于区块链的数字化古籍管理模型智能合约协议算法

基于区块链的数字化古籍管理模型是通过建立智能合约,自动执行实现数字化古籍的版权登记、使用交易授权。

4.3.1 数字化古籍版权登记智能合约 古籍是古人留下的著作,原始作者已经去世几百上千年,其版权不是谁独有,但古籍的传承保护靠的是收藏单位、个人,以及对古籍进行整理、出版工作的单位、学者、出版社等,他们都有相应的权益。形成的数字化古籍全文影像数据、图文对照数据、文本数据等享有相应的版权。数字化古籍管理区块链模型建立的版权登记智能合约(Copyright Registration Contract,CRC)按照“谁先创作、谁先申请、谁就拥有该项版权”的原则。用户节点在安装客户端注册的时候,会同时在区块链上部署一份CRC,通过初始化注册函数构造合约拥有者。如果有节点调用 CRC 时,该合约会先检测交易发起方的信息。如果合约的调用者不是这份合约的拥有者,那么则无法将该古籍版权写入区块链。如果数字化古籍版权登记成功,智能合约 CRC 会返回由该数据特征值提取的 Hash 值,作为这一数字化古籍版权的标识[20]。智能合约 CRC 算法如算法 3 所示:

算法 3:数字化古籍版权登记合约

Input: newblockB, blockchainC, T_x), transaction object

Output: if error, registration failure else return ID

```
procedure register(Tx)
    if msg. sender = owner then
      Tx = new Transaction();
      Tx. name = Tx. name;
      Tx. time = now;
      Tx. content = Tx. content;
      hash = createID()
      blockchainC←l
      return hash
    end if
end procedure
```

4.3.2 数字化古籍版权授权智能合约 数字化古籍管理区块链能使古籍所有者对古籍版权的管理具有灵活性,按照版权授权合约(Copyright License Contract, CLC),当有用户向版权用户发出使用授权请求时,版权用户在接到使用请求后,会调用智能合约 CLC 中的古籍版权使用授权协议,触发本次交易。版权用户如果同意本次交易事件,智能合约 CLC 会调用 transfer 函数执行本次授权交易。如果版权用户不同意该用户使用自己的数据,则从交易队列中停止本次交易。智能合约 CLC 算法如算法 4 所示:

算法 4:数字化古籍版权使用授权合约

Input: T_x, transaction object

```
procedure receiveTx(tx)
    if msg. sender = owner then
      if tx. operation = = true then
        transaction(tx. to, tx. from)
        digitalWork = getCIC(tx, hash);
        removeFrom(owner, digitalWork);
        linkTo(tx. requester, digitalWork);
      end if
      stop(tx)
    end if
end procedure
```

5 数字化古籍分布式存储及激励机制

如果说通过部署区块链智能合约相应的协议算法,可以从版权保护、流通使用方面解决数字化古籍管理问题。那么,区块链的分布式存储与共识机制,可以解决数字化古籍存储与传统古籍数据库激励机制匮乏问题。

5.1 数字化古籍的分布式存储

目前数字化古籍数据都是采用中心化的数据库存储,随着数据量增加而需要大量存储空间,而将这些古籍数据从中央数据库分发采用 HTTP 协议,当内容过度集中化之后,会导致高度依赖 Internet 骨干网,虽然有利于管理,但会使效率降低或者出现路由表失控,产生数据安全等一系列问题。

从 4.2 节可知,区块链上区块只记录了交易信息数据、智能合约、一些摘要备注等信息和文件大小小于256K 的数据,这是由区块链定义的区块结构所决定的,从而避免了较大数据直接存储到区块链上而导致

存储暴增。区块链技术为解决这类大数据的存储，采用的是IPFS(The Inter Planetary File System，星际文件存储系统)是一个P2P的分布式数据分发协议，能够将所有具有相同文件管理模式的计算设备连接在一起。因此我们构建数字化古籍管理区块链模型从根本上改变了数字化古籍的存储、搜索、下载方式[21]。

IPFS对存储各类文件是通用的，存储的文件没有大小的限制，而诸如全影像和图文对照等古籍数据，单个数据文件就很大。IPFS系统会自动将大的数据文件拆分区划为多个小数据块文件，然后再把小数据块文件分发存储到众多节点电脑上。同时为了保证数据的安全性和可靠性，区块链IPFS分布式存储协议会将每份数字化古籍数据复制3份以上，分散存储。区块链用户节点可以分为两类，一类是轻节点，即为普通消费用户，能够构造区块、完成智能合约、验证区块等，实现数字化古籍区块链维护、版权登记、查询、下载等操作，下载客户端上线即可；另一类是全节点或者称为IPFS节点，除了能够进行轻节点的所有操作外，还贡献出一定的存储空间用来存储数据，实现"人人为我，我为人人"的共享理念。在区块链共识机制作用下的激励机制也更多地倾向于全节点用户。

数字化古籍被上传到IPFS节点后，将生成一个新名字，这个名字实际上是根据数据内容计算出的一个加密Hash值。加密能够保证该Hash值始终只表示这一数据的内容，哪怕只在数据中修改一个比特的数据，Hash值都会完全不同。同时IPFS系统会计算全网IPFS节点数量，将数字化古籍及副本数分成若干数据块存储于各个IPFS节点。

另外，IPFS从根本上改变了节点用户的搜索和下载方式，当节点通过智能合约发出搜索请求时，IPFS通过使用一个分布式Hash表查找对应Hash值，可以快速(在一个拥有10,000,000个节点的网络中只需20跳)检索到拥有该古籍数据的节点，并使用Hash验证其是否是正确的数据，使用户能够快速搜索到自己想要的内容。如果下载会通过使用授权智能合约进行类似BitTorrent方式进行下载。通过智能合约授权后下载用户会被提供一个种子(seed)文件，种子文件中提供计算出的拥有该古籍数据块节点地址，同时提供P2P下载协议进行快速下载。

5.2 数字化古籍管理区块链网络激励机制

共识机制是区块链的基石，它相当于维系区块链网络正常运转的法律，是参与者在区块链网络中贡献以及获取权益的证明。数字化古籍区块链网络通过建立智能合约，然后形成全网节点共识，创造出区块链上Trustless(免信任的)记账机构，保证数字化古籍的每次上传、下载、检索、浏览等所有操作在所有记账节点上的一致性，参与者谁的贡献大谁获得的激励就越多。区块链经济模式的"共识机制"和"去中心化"是用代币(Token)来激活内部生态，从而建立一个去中心化自组织数字化古籍管理体系。Token是附着于区块链网络并在该网络内产生和使用的记账单位。Token是一个激励机制的基础，是一个区块链生态里的一种权益证明。

通常情况下Token一经发行，便严格按照区块链代码执行，不受个人或机构控制。在这里我们假设构建的数字化古籍管理区块链网络发展良好，形成一个古籍管理、保护、开发、利用的和谐的区块链生态系统。那么可以根据其未来成长预期设定Token的发行总量，永不增发建立一套通缩的代币，本研究借鉴国外基于区块链的内容型社区Steem和国内区块链知识内容社区币乎的激励机制，对数字化古籍管理区块链网络激励机制进行相应的探讨。例如：①数字化古籍资源建设池(占Token总量50%)每年释放所属池余额的10%，奖励给提供数字化古籍的版权用户，多劳多得；②数字化古籍价值奖励池(占30%)每年释放所属池余额的10%，按数字化古籍利用率奖励给版权拥有者；③数据存储安全维护池(占15%)每年释放所属池余额的10%，奖励给提供存储空间的IPFS节点用户；④数字化古籍区块链网络运行和推广池(5%)分配给制定标准、管理、维护中心运行推广者，按比例逐年行权。

Token的价值体现应该是多方面的：一方面，古籍自身价值毋庸置疑，是传承文化的载体。Token的多少体现了为古籍文献保护做出贡献的大小，可以作为国家奖励机制的标准。另一方面，市场价值，持有Token相当于拥有数字化古籍管理网络系统的股权，会随着古籍产业价值提升而提升。

6 结语

区块链作为颠覆互联网的创新技术，逐渐进入高速发展阶段，政府持续出台相关文件鼓励区块链技术的应用发展，国内外各领域也开始逐渐布局区块链+。众多区块链项目落地并且发展良好，如版权领域的Blockai、Ownership、人人链、纸贵科技等，防伪溯源领域的Chronicled、Vechain、阿里、京东防伪追溯平台等，数据安全存储领域的Factom、迅雷链等。2018年4月百

度区块链原创图片服务平台“百度图腾”上线，采用区块链技术进行版权登记，通过时间戳为每张原创图片生成版权 DNA，从而实现图片生产、版权存证、图片存储、共享分发、交易变现、侵权监测、维权服务等全链路服务平台。

本文提出了数字化古籍管理区块链体系模型，针对当前古籍版权保护利用存在的保护难、举证难、维权难等问题。模型设计了区块链古籍交易信息存储机制，通过动态交易数据存储体系，交易信息数据经过节点验证组合验证形成全网共识，使交易信息动态增加；建立智能合约算法协议，自动化执行且不可干预的机制保证了模型运行的高效透明；IPFS 系统 P2P 分布式存储实现数字化古籍数据的安全存储、共识机制建立合理的激励机制。这些相结合实现了数字化古籍的有效管理。本文只是对区块链技术解决数字化古籍管理问题做了初步的研究，要具体实施还会存在一些问题，诸如：底层区块链平台选择、客户端开发、智能合约编写等问题。此外一个区块链项目成功与否最重要的是智能合约的构建及执行，智能合约代码编写是否合理高效，是否符合逻辑都是关键问题。目前，区块链应用还处于早期发展阶段，但不可否认区块链技术自身特性决定其在知识产权、数据保护、档案管理、出版行业等众多领域有不可估量的应用价值。

参考文献：

[1] 曹天晓. 新技术下古籍数字化分类及意义探究[J]. 图书馆研究与工作，2017(9)：37－41.

[2] 聂灿. 千年古籍讲述纸张上文明[N]. 深圳商报，2018－05－24(B5).

[3] 黄水清，王东波. 古文信息处理研究的现状及趋势[J]. 图书情报工作，2017，61(12)：43－49.

[4] 余力，管家娃. 我国古籍数字化建设现状分析及发展研究[J]. 数字图书馆论坛，2017(11)：41－47.

[5] 申屠晓明. 传媒行业区块链应用模式与技术方案解析[J]. 传媒评论，2018(4)：27－31.

[6] 贾引狮. 基于区块链技术的网络版权交易问题研究[J]. 科技与出版，2018(7)：90－98.

[7] 郑阳，杜荣. 区块链技术在数字知识资产管理中的应用[J]. 出版科学，2018(3)：97－104.

[8] 李悦，黄俊钦，王瑞锦. 基于区块链的数字作品 DCI 管控模型[J]. 计算机应用，2017，37(11)：3281－3287.

[9] 郝琨，信俊昌，黄达，等. 去中心化的分布式存储模型[J]. 计算机工程与应用，2017，53(24)：1－7，22.

[10] 吴振铨，梁宇辉，康嘉文，等. 基于联盟区块链的智能电网数据安全存储与共享系统[J]. 计算机应用，2017，37(10)：2742－2747.

[11] 王继业，高灵超，董爱强，等. 基于区块链的数据安全共享网络体系研究[J]. 计算机研究与发展，2017，54(4)：742－749.

[12] GRINBERG R. Bitcoin: an innovative alternative digital currency [EB/OL]. [2017－03－01]. https://www.researchgate.net/publication/228199328_Bitcoin_An_Innovative_Alternative_Digital_Currency.

[13] 孟奇勋，吴乙婕. 区块链视角下网络著作权交易的技术之道[J]. 出版科学，2017，25(6)：25－31.

[14] 袁勇，周涛，周傲英，等. 区块链技术：从数据智能到知识自动化[J]. 自动化学报，2017，43(9)：1485－1490.

[15] SZABO N. Formalizing and securing relationships on public networks[EB/OL]. [2016－11－20]. http://journals.uic.edu/ojs/index.php/fm/article/view/548/469.

[16] 王帅宇，李晨. 一种基于区块链技术的大数据确权方法及系统：中国，CN106815728A [P]. 2017－06－09. http://nvsm.cnki.net/kns/detail/detail.aspx? QueryID = 5&CurRec = 1&dbcode = SCPD&dbname = SCPD2017&filename = CN106815728A.

[17] ZHAO J L，FAN S，YAN J. Overview of business innovations and research opportunities in blockchain and introduction to the special issue[J]. Financial innovation，2016，2(28)：2－7.

[18] 乔蕊，董仕，魏强，等. 基于区块链技术的动态数据存储安全机制研究[J]. 计算机科学，2018，45(2)：57－62.

[19] GRAMOLI V. From blockchain consensus back to Byzantine consensus [J]. Future generation computer systems，2017，9(23)：1－20.

[20] KOSBA A，MILLER A，SHI E，et al. Hawk: the blockchain model of cryptography and privacy-preserving smart contracts[C]//Proceedings of the 2016 IEEE symposium on security and privacy. Piscataway: IEEE，2016：839－858.

[21] 殷龙，王宏伟. 基于 IPFS 的分布式数据共享系统的研究[J]. 物联网技术，2016，6(6)：60－62.

作者贡献说明：

高利：提出研究思路，撰写论文；

王春艳：提出修改意见；

高心丹：指导研究思路，设计论文框架。

Research on Construction of Digital Ancient Book Management System Model Based on Block Chain Technology

Gao Li[1] Wang Chunyan[1] Gao Xindan[2]

[1] Northeast Forestry University Library, Harbin 150040

[2] Northeast Forestry University College of Information and Computer Engineering, Harbin 150040

Abstract: [**Purpose/significance**] From the perspective of blockchain technology, this paper proposes a solution to the problems of the protection and use of ancient books, and proposes a solution to the digital block management system model for ancient books, and implements an ancient management innovation model. [**Method/process**] First of all, a block chain decentralized P2P architecture is designed from the aspects of construction and implementation. The entire network consensus link block is realized by the node initiating the transaction, the construction block, and the validity verification. Secondly, an information storage mechanism and smart contract protocol algorithm is estabalished to automatically implement the guarantee model in an efficient and transparent manner, and the digitalized ancient books copyright registration and circulation use information is realized, which cannot be tampered with and traceable. Finally, the consensus mechanism and the distributed storage mechanism of the IPFS protocol system solve scientific data storage and security, and at the same time establish a reasonable incentive mechanism to achieve effective management of scientific data. [**Result/conclusion**] Through the establishment of this model, the traditional centralized management system of ancient books database is abolished, and the entire process of protection and utilization of ancient books can be traced and cannot be tampered with the form of block + chain to achieve the technological innovation of protection of ancient books.

Keywords: blockchain digital classics smart contracts IPFS

2019’《图书情报工作》优秀论文

本刊自 2014 年起发布当年及前两年高被引论文 TOP10。自 2019 年起,《图书情报工作》将评选优秀论文,以中国知网检索到的前 5 年发表的高被引和高下载论文各年 TOP50 为基础,兼顾发表时间,由编辑部最后选定优秀论文 20 篇进行公布(见下表)。被选定的优秀论文第一作者将获得由《图书情报工作》杂志社颁发的优秀论文证书,并赠送全年期刊论文电子版。

序号	题名	作者	发表时间
1	基于动态 LDA 主题模型的内容主题挖掘与演化	胡吉明　陈果	2014,58(2)
2	新媒体技术发展对网络舆情信息工作的影响研究	魏超	2014,58(1)
3	基于 LDA 模型和微博热度的热点挖掘	唐晓波　向坤	2014,58(5)
4	大数据环境下多源信息融合的理论与应用探讨	化柏林　李广建	2015,59(16)
5	Google 三大云计算技术对海量数据分析流程的技术改进优化研究	卢小宾　王涛	2015,59 (3)
6	美国高校信息素养标准的改进与启示——ACRL《高等教育信息素养框架》解读	秦小燕	2015,59(19)
7	高校图书馆微信公众平台传播影响力评价体系研究	郭顺利　张向先　李中梅	2016,60(4)
8	微信用户信息共享行为影响因素模型及实证研究——基于信息生态视角的分析	王晰巍　曹茹烨　杨梦晴等	2016,60(15)
9	数字学术中心:图书馆服务转型与空间变革——以北美地区大学图书馆为例	介凤　盛兴军	2016,60(13)
10	开放政府数据评估框架、指标与方法研究	郑磊　关文雯	2016,60(18)
11	不同语料下基于 LDA 主题模型的科学文献主题抽取效果分析	关鹏　王曰芬　傅柱	2016,60(2)
12	高校图书馆参与高校智库建设与服务的优势及路径研究	赵雪岩　彭焱	2016,60(22)
13	用户在线知识付费行为影响因素研究	张帅　王文韬　李晶	2017,61(10)
14	基于用户行为建模和大数据挖掘的图书馆个性化服务研究	何胜　冯新翎　武群辉等	2017,61(1)
15	“双一流”建设背景下高校图书馆服务 ESI 学科建设的内容与策略	刘勇	2017,61(9)
16	我国政府数据开放共享的政策框架与内容:国家层面政策文本的内容分析	黄如花　温芳芳	2017,61(20)
17	图书馆阅读推广的合理性审视	范并思	2017,61(23)
18	数字学术环境下学术图书馆发展新形态研究——以空间、资源和服务“三要素”为视角	刘兹恒　涂志芳	2017,61(16)
19	图书馆与智库	初景利　唐果媛	2018,62(1)
20	基于信息熵的新媒体环境下网络节点影响力研究——以微信公众号为例	邢云菲　王晰巍　韩雪雯等	2018,62(5)